DOCUMENT RESUME

ED 127 372 95 TM 005 532

AUTHOR Wheeler, Patricia; Elias, Patricia J.
TITLE Beginning Teacher Evaluation Study: Phase II, 1973-74, Final Report: Volume V.3. Historical Data Study.
INSTITUTION California State Commission for Teacher Preparation and Licensing, Sacramento.; Educational Testing Service, Princeton, N.J.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
REPORT NO ETS-PR-76-14
PUB DATE Jul 76
NOTE 25p.; For related documents, see TM 005 524-535

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Academic Achievement; Achievement Tests; *Data Collection; Effective Teaching; Elementary Education; Grade 2; Grade 5; Mathematics; Nonparametric Statistics; Reading; Standardized Tests; *Test Results; *Time

ABSTRACT

The Beginning Teacher Evaluation Study (BTES), Phase II, was a research project on effective teaching behavior--what teachers do that significantly affects what and how pupils learn. The purposes of Phase II were to (1) develop an assessment system for measuring teacher and pupil behaviors and other factors which could influence each of them and their interrelationships and (2) generate hypotheses about the interrelationships among teacher and pupil behaviors and related factors. Forty-one second grade and 54 fifth grade experienced teachers participated in the study. The purpose of this volume of the final report was to assess the stability of teacher output in terms of student achievement during and prior to Phase II of the study. The procedure involved the collection of achievement data, specifically skills on norm-referenced tests for students in the classrooms of the sample teachers for the previous two years. The limited historical data available from BTES classrooms and schools suggest that it is inappropriate to formulate any hypotheses about the stability of past teacher performance over time in terms of accessible information about pupil achievement. (RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED127372

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
DOROTHY URBAN
ETS
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER.

TM005 532

PR-76-14

BEGINNING TEACHER EVALUATION STUDY
PHASE II 1973-74

FINAL REPORT

VOLUME V.3.

HISTORICAL DATA STUDY

BY

PATRICIA WHEELER
PATRICIA ELIAS

EDUCATIONAL TESTING SERVICE
PRINCETON, NEW JERSEY

BEGINNING TEACHER EVALUATION STUDY: PHASE II

FINAL REPORT: VOLUME V.3.

HISTORICAL DATA STUDY

by

Patricia Wheeler

Patricia Elias

Educational Testing Service
Berkeley, California

A project conducted by Educational Testing Service for the California Commission for Teacher Preparation and Licensing and funded by the National Institute of Education.

Copyright © 1975, 1976 by Educational Testing Service. All rights reserved.

PREFACE

The Beginning Teacher Evaluation Study (BTES) is a long-term project of the California Commission for Teacher Preparation and Licensing. The Commission is responsible for licensing teachers in California and is trying to determine what factors should be considered in this process.

The second phase of the study was conducted by Educational Testing Service for the Commission. Phase II was the hypotheses-generating and instrument-development phase of BTES. ETS had two tasks: (1) to develop an assessment system to measure both teacher and pupil behaviors as well as other factors which might be related to these behaviors; and (2) to generate hypotheses about the interrelationships between teacher and pupil behaviors and related factors.

The study was conducted in 43 schools in eight districts throughout the state of California. A total of 41 second grade teachers and 54 fifth grade teachers participated in the project during Phase II.

The final report for Phase II consists of several volumes. Volume I describes the design and rationale for the experimental design and data analysis procedures and includes the major findings of Phase II. Volume II describes the conduct of the field study and the sample of participants.

Because of the complex nature of Phase II, a variety of techniques was used to measure teacher and pupil behaviors. They are described in Volumes III, IV, and V. Results are also included in these volumes.

Volume III describes the observation systems in detail and is available in three separately bound sections. The first section, Volume III.1., describes the behavior recording observation system used in the project--APPLE (Anectdotal Process for Promoting the Learning Experience). Volume III.2. describes the category system used to observe classroom activities--

RAMOS (Reading and Mathematics Observation System). The third section of this volume, III.3., covers the videotaping of instructional activities during reading and mathematics.

Volume IV concerns other aspects of the measurement system and covers both the pupil and teacher test batteries.

The fifth volume covers a series of small studies done as part of Phase II. Volume V.1. looks at teacher aptitudes as related to teacher behaviors. Volume V.2. is concerned with the relationship between teacher expectations and pupil performance. Volume V.3. reviews performance of pupils in the BTES teachers' classrooms for two years prior to Phase II, the historical test data. Volume V.4. discusses the Diagnostic Film Test, a device designed to assess teachers' skills in diagnosing reading problems and prescribing corrective action. Volume V.5. summarizes the results of work diaries completed by the teachers on their reading and mathematics instructional program.

Information on the availability of these volumes can be obtained from:

Dr. Frederick J. McDonald
Educational Studies
Educational Testing Service
Princeton, NJ 08540

Information on other phases of BTES can be obtained from:

California Commission for Teacher
Preparation and Licensing
1020 O Street
Sacramento, CA 95814

ACKNOWLEDGEMENTS

The authors wish to acknowledge the assistance of Mary Corder and Gail Hare of the Berkeley Office of Educational Testing Service and J. Richard Harsh of ETS's Los Angeles Office in the collection and analysis of test data and the preparation of this report. The cooperation of school personnel in the collection of test data is also appreciated.

TABLE OF CONTENTS

	Page
Preface	i
Acknowledgements	iii
Historical Data Study	1
Collection Procedures	1
Description of Historical Data Available	2
Analysis Procedures and Results	4
Limitations to the Interpretation of the Historical Data	9
Conclusions	11
References	13
Appendix	15

LIST OF TABLES

Table		Page
1	Correlations of Mean Pupil Outcome in Reading, Grade 2	5
2	Number of Teachers in Each Category by Geographic Location and SES Level of Their School Site	7
3	Correlations of Mean Pupil Outcomes in Reading, Grade 2, by Geographic Area, Spring 1973-Spring 1974	8
4	Correlations of Mean Pupil Outcomes in Reading, Grade 2, by Socioeconomic Level, Spring 1973-Spring 1974	8
	APPENDIX:	
5	Historical Pupil Achievement Data Available for BTES Second Grade Teachers	17
6	Historical Pupil Achievement Data Available for BTES Fifth Grade Teachers	18

LIST OF FIGURES

Figure		Page
1	Scatter diagram of reading outcome measures	6

Historical Data Study

Phase II of the Beginning Teacher Evaluation Study (BTES) was devoted to the generation of hypotheses and development of instruments to assess the impact of teacher skills and teacher performance on pupil achievement in reading and mathematics. The purpose of the stability substudy, as stated in an addendum to the Phase II proposal, was as follows:

> "The stability substudy will assess the stability of teacher output in terms of student achievement during and prior to the phase of the study. The procedure will involve the collection of achievement data, specifically skills on norm-referenced tests for students in the classrooms of the sample teachers for the previous two years."

Thus the task in this substudy was to collect historical data on the achievement of pupils in the classroom of BTES teachers in previous years, and to compare this historical data with data gathered through the administration of pupil outcome measures during Phase II.

Collection Procedures

In order to accomplish this task, ETS staff collected available standardized achievement test data in reading and mathematics for students who were in the classroom of the BTES teachers during the 1971-72 and 1972-73 school years. The collection was done between April and October of 1974, and was scheduled at times convenient to the participating districts.

Preliminary arrangements for the collection of the historical data were made in January and February of 1974, when BTES Associate Project Directors met with each of the principals in the 43 project schools to review the progress of the study to that time. The principals were

advised of the historical data component of the research design, their sanction was obtained for gathering this information, and data collection procedures were established for each school and district.

The data collection on achievement of pupils in the BTES classrooms during the project year (1973-74) is described in detail in Volume II of this report. The norm-referenced student outcome measures for Phase II of the study were reading and mathematics sections of the California Achievement Tests, 1970 Edition (CTB/McGraw-Hill).

Description of Historical Data Available

A variety of standardized (norm-referenced) tests were used in the BTES teachers' classrooms in 1971-72 and 1972-73. There were district to district, school to school and classroom to classroom differences among the tests that were administered. In some cases, test scores were obtained for both reading and mathematics; in others, scores were available for only one subject area. Scores were reported as raw scores, grade equivalent scores, percentile ranks, national percentile ranks, scaled scores, stanines and national stanines. The tests administered included the following:

Comprehensive Tests of Basic Skills, CTB/McGraw Hill
Cooperative Primary Tests, Educational Testing Service
Iowa Tests of Basic Skills. Houghton Mifflin
Iowa Tests of Educational Development, Science Research Associates (SRA)
SRA Achievement Series, SRA
SRA Modern Math Understanding Test, SRA
Stanford Diagnostic Reading Test, Harcourt Brace Javonovich

District testing programs in California have frequently been built around the state mandated testing programs. There was no state mandated program at the fifth grade level during 1971-72 or 1972-73. Consequently, in some districts no standardized tests were administered at that grade

level. A state mandated test of reading achievement for grade two was included in the state testing program in California during those two years. Administration of the Cooperative Primary Reading Test, Form 23A (Educational Testing Service), was required in the spring of the school year (May). No standardized tests were required in mathematics at the second grade level. (See California State Testing Program, 1971-72 and 1972-73, Profiles of School District Performance for a description of the state mandated testing program in California.)

Several schools had previously administered standardized tests during the fall of grade five. However, since these tests essentially were measures of skills pupils brought to the classrooms and were administered before the BTES teachers could have had any impact on these pupils, the results were not included in the analysis of the historical data.

Unless a school had a special program such as those funded under Title I or Title VII, standardized tests were not routinely administered during both the fall and spring of the school year. Where post-test (output) data were available, there were usually no pre-test (input) data, and vice versa. There were, in fact, only three classrooms in which both sets of data were obtainable.

Test data were available from seven of the eight participating districts. The eighth district was unable to furnish any data for 1971-72 and 1972-73 because of computer failures during those years. Some kind of retrospective information on pupil performance was available for 52 of the 95 BTES teachers: 32 of the 41 second grade teachers and 20 of the 54 fifth grade teachers.

The only case where sufficient test data were available for comparison was in grade two for the Cooperative Primary Reading Test (CPRT) for the spring of 1972 and spring of 1973. After eliminating second grade teachers with test data who currently or previously were members of a teaching team, had taught at a different grade level in previous years, or were not responsible for teaching reading to pupils in their classrooms, a total of 24 teachers were included in the sample for the stability substudy. The variety of tests used at the fifth grade level and the lack of output measures for the classrooms made it inappropriate to do an analysis of historical test data at that level.

The Appendix gives some notion of the variety of historical data collected. From that range of data, a total of only 24 second grade teachers in five districts in California could be included in the analysis. The districts included were Los Angeles, Riverside, San Jose, San Marino, and Vacaville. Only information on reading outcomes was available for analysis.

Analysis Procedures and Results

Nonparametric techniques were employed in the analysis of the historical test data. Statewide information was not available on the total population of California teachers who had taught the same grade levels from 1971-72 through 1973-74. There was, therefore, no way to determine if the stability substudy sample was representative of the total California population.

For purposes of analysis of the CPRT (Form 23A) test data, all scores were converted to raw scores if they were not already in that format. The mean scores on the California Achievement Tests: Reading Comprehension (CAT:RC) and the CPRT were computed for each second grade

teacher in the study. If a teacher had fewer than eight pupils with reading scores for a given year, the data for that year were excluded from the analysis. We felt that such small groups might not be representative of pupil outcome levels in reading for that teacher.

The Spearman rank order correlation (Siegel, 1956) was computed to measure the stability of pupil outcomes in reading. Comparisons were made between Spring 1973-Spring 1974, Spring 1972-Spring 1974, and Spring 1972-Spring 1973 mean scores. Spring 1972 test data were available for ten teachers in three districts. Twenty-four teachers in five districts were included in the Spring 1973-Spring 1974 comparisons. The results are given in Table 1.

Table 1

Correlations of Mean Pupil Outcome in Reading, Grade 2

Years	r_s	N
1973-1974	+.61*	24
1972-1974	+.84*	10
1972-1973	+.94*	10
1973-1974	+.82*	10

*Significant at the .01 level.

Two comparisons were done for the 1973-1974 test data. One calculation was based on the 24 teachers who had test data available for both of those years. The second comparison was based on the ten teachers who had test data available across all three years. The ten teachers appear to have had more stable pupil reading outcomes than the other 14

teachers included in the larger analysis, as indicated by the levels of positive correlations (+.61 compared to +.82). The CPRT only was used in 1972 and 1973, which probably makes the correlation for these two years higher (+.94) then for 1972-1974 and 1973-1974, where the CPRT was correlated with the CAT:RC.

Figure 1 shows a scatter diagram of Spring 1973 CPRT mean scores plotted against Spring 1974 CAT:RC mean scores. The positive relation between these two sets of data is apparent.

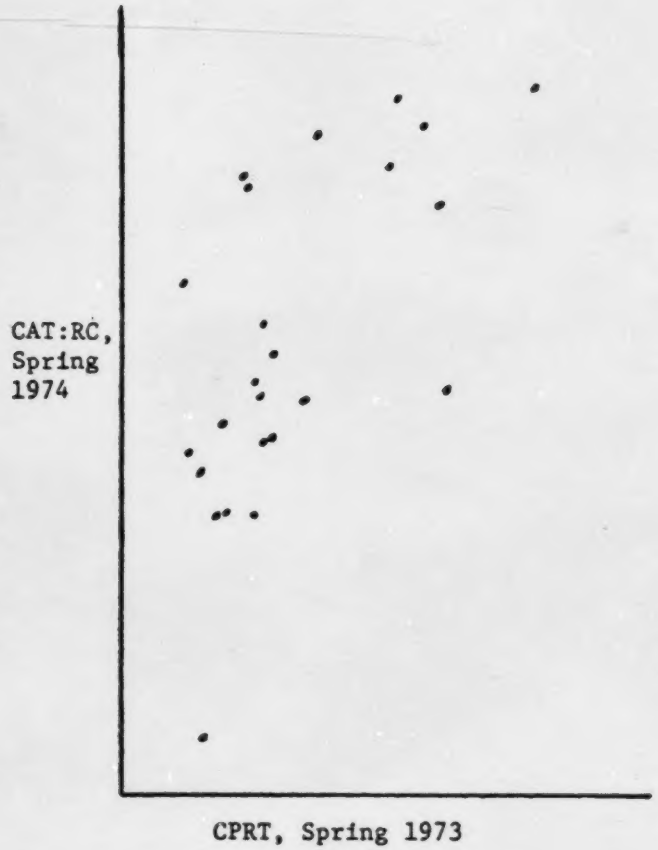

Figure 1. Scatter diagram of reading outcome measures.

The 1973-1974 sets of data were grouped and analyzed by both geographic area and by socioeconomic level of the school, as judged by the principal in the Principal Interview (conducted in the Spring of 1974 as part of the total study). Table 2 shows the categorization of the historical data by geographic location and socioeconomic status (SES) levels. The classrooms represent a variety of SES levels and geographic locations.

Table 2

Number of Teachers in Each Category
by Geographic Location and SES Level
of Their School Site

Geographic Location	Socioeconomic Level			Total
	Low-Low Middle	Middle	High-High Middle	
Urban	6	4	2	12
Suburban	2	0	5	7
Rural	2	2	1	5
Total	10	6	8	

The correlations between CPRT mean scores for Spring 1973 and the CAT:RC mean scores for Spring 1974 in terms of the two sets of demographic indices are shown on Tables 3 and 4.

Table 3

Correlations of Mean Pupil Outcomes in Reading, Grade 2, by Geographic Area, Spring 1973-Spring 1974

Type of Area	r_s	N
Urban	+.06	12
Suburban	+.71*	7
Rural	+.80	5

*Significant at the .05 level.

Table 4

Correlations of Mean Pupil Outcomes in Reading, Grade 2, by Socioeconomic Level, Spring 1973-Spring 1974

SES Level	r_s	N
Low-Low Middle	+.07	10
Middle	+.77	6
High-High Middle	+.38	8

Limitations to the Interpretation of the Historical Data

The correlations in Table 1 indicate that there is some stability from year to year in reading outcomes for pupils in the 24 BTES teachers' classrooms. All the correlations are significant at the .01 level. Correlations at this level could be obtained by chance less than one percent of the time. However, one must exercise caution in concluding from this that teachers are equally effective in teaching reading to their pupils from one year to the next. There are several limitations to the data collected for the stability substudy of BTES.

The historical data are not unlike the kinds of information available to administrators and others in decision making roles about teaching performance. The kinds of constraints inherent in analyzing and interpreting historical data are equally important for researchers and administrators. Information which might have allowed the description of the achievement or aptitude characteristics of the pupils whose test scores were reported was either inaccessible or nonexistant. Pre- and post-test achievement scores were available for only three classrooms. Any interpretations about the effectiveness of BTES teachers with their previous groups of pupils should be made with the understanding that the interpretations are based on pupils for whom information on entry level skills in reading and mathematics was unavailable.

The analysis of the historical data is based on the assumption that pupils for a given teacher are approximately the same from year to year. In those areas where one might expect greater variation in the pupil population, we obtained lower correlations--Urban (+.06) and Low-Low Middle SES (+.07). (See Tables 3 and 4.) Such low correlations imply considerable variability from year to year. Pupils are not necessarily

assigned randomly to classrooms. While teachers may have had heterogeneous groups of pupils each successive year, they also may have had relatively slow or fast groups of children in one year or another. Year-to-year variations in pupil outcomes might reflect more about differences in pupil population homogeneity and school practice in assigning pupils to classrooms than they do about the stability of teacher performance.

When the historical data were organized by the socioeconomic level of the school site and by the type of geographical area, the only significant correlation was for teachers in the suburban schools. Suburban schools might be the most homogeneous in terms of pupil characteristics and the most stable in terms of pupil mobility. Without more information, however, we cannot determine if this is the case.

The interpretation difficulties are further confounded by the unresolved question of the accuracy of the historical data. In some cases, the CPRT had been scored by a professional scoring service while in others the tests had been scored by the district staff or school staff. No information was available on the accuracy of the scoring or the quality control procedures employed. We had to assume the scores were accurate as reported for purposes of analysis in the stability substudy. The amount of cumulative error cannot be determined without more information, much of which is no longer available (e.g., actual pupil test booklets to permit rescoring).

Another obvious limitation in interpretation was the assumption that the BTES teachers actually taught reading and mathematics to all of the pupils during any given year. Information was available only for 1973-1974, the project year. It is not uncommon for some or all of the

children in a classroom to go to other teachers or to specialists for instruction in reading and mathematics.

A general methodological limitation reflects problems in using classroom averages as the single index of teacher performance. The mean of all pupils' performance may be a meagre and inaccurate characterization of a teacher's strategies and skills. A teacher who concentrated on the least able pupils in a class might have a less impressive overall classroom average than a teacher who worked essentially with the most able children. Mean scores are affected by extreme scores.

Conclusions

For this hypothesis-generating phase of the BTES, the focus of the stability substudy was to examine existing historical pupil achievement data in reading and mathematics from BTES teachers' classrooms for previous years and to determine whether those data could be used in conjunction with the project data collected in 1973-1974 to evaluate the stability of teacher performance over a three year period. Surprisingly little standardized achievement test data was routinely collected in basic academic areas in the second and fifth grade classrooms of BTES schools. Post-test or output information appeared to have been collected primarily when it was mandated by the California State Department of Education. It was not mandated at the fifth grade level. Pre-test or pupil input information was either not available or was inaccessible one or two years after having been collected. Any interpretations about the effectiveness of BTES teachers with their previous pupils or about the stability of impact of BTES teachers on pupil outcomes over a period of three years must be made with the understanding that the results are based on pupils of unknown aptitude and entry level

skills in reading and mathematics.

Based on the data that were analyzed, correlations between the years of historical information about pupil performances suggested that the pupil performance in the classroom of a given teacher tended to be stable over time. However, when the data were organized in ways that began to reflect something about the characteristics of the pupil populations, the correlations suggested that year-to-year relationships among performances varied substantially by type of geographic location and by SES level of schools. It was not possible to tell from the types of information available whether these variations were due to pupil differences, teacher differences, or both, but it is reasonable to assume that the demographic indices represent pupil population differences to some extent. Even with the relatively crude demographic indices we used, the notion of stability of teacher performances as reflected solely by pupil outcome measures began to evidence serious limitations.

The kinds of limitations inherent in historical pupil outcome data and the analysis and interpretation of those kinds of data are important considerations for both administrators and educational researchers. The analysis of the data utilizing fairly rough demographic indices of school characteristics as a proxy for pupil characteristics illustrated the fallacy in thinking about teacher performance in terms of the single available indicator of pupil outcome. To test the stability of teacher performance over time, researchers and administrators at an absolute minimum must collect and analyze both input and output data on pupils. The limited historical data available from BTES classrooms and schools suggest that it is inappropriate to formulate any hypotheses about the stability of past teacher performance over time in terms of accessible information about pupil achievement.

References

California State Department of Education. California state testing program, 1971-72 and 1972-73, profiles of school district performance. Sacramento: California State Department of Education, 1974.

Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

APPENDIX

Table 5

Historical Pupil Achievement Data Available for BTES Second Grade Teachers

Number of Teachers	Grade Taught at the Time of Testing	Dates of Testing	Test	Subject Area Tested	
				Reading	Math
14	2	5/73	Cooperative Primary Tests: 23A	X	
1	2	5/73	Cooperative Primary Tests: 23A	X	X
5	2	5/72	Cooperative Primary Tests: 23A	X	
		5/73	Cooperative Primary Tests: 23A	X	
1	2	5/71	Cooperative Primary Tests: 23A	X	
		5/72	Cooperative Primary Tests: 23A	X	
2	2	5/72	Cooperative Primary Tests: 23A	X	X
		5/73	Cooperative Primary Tests: 23A	X	X
2	2	5/72	Cooperative Primary Tests: 23A	X	
		5/73	Cooperative Primary Tests: 23A	X	
		4/73	SRA Achievement Series: F	X	X
1	2	5/72	Cooperative Primary Tests: 23A	X	
		5/73	Cooperative Primary Tests: 23A	X	
		10/72-4/73*	SRA Achievement Series: F	X	X
1	2	5/72	Cooperative Primary Tests: 23A	X	
		5/73	Cooperative Primary Tests: 23A	X	
		5/71	SRA Achievement Series: C	X	
		4/72	SRA Achievement Series: C	X	
		5/71	Modern Math Understanding: Primary C		X
		4/72	Modern Math Understanding: Primary C		X
2**	2/3	5/73	Cooperative Primary Tests: 23A	X	X
			Cooperative Primary Tests: 23B	X	X
1	2/3	5/72	Cooperative Primary Tests: 23A	X	X
			Cooperative Primary Tests: 23B	X	X
1	3	5/73	Cooperative Primary Tests: 23B	X	
1	4	10/72	Comprehensive Test of Basic Skills: 2Q	X	X

N=32

* Pre- and Post tests
** Team teaching

Table 6

Historical Pupil Achievement Data Available for BTES Fifth Grade Teachers

Number of Teachers	Grade Taught at the Time of Testing	Dates of Testing	Test	Subject Area Tested: Reading	Subject Area Tested: Math
1	2/3	5/72	Cooperative Primary Tests: 23A	X	X
			Cooperative Primary Tests: 23B	X	X
		5/73	Cooperative Primary Tests: 23A	X	X
			Cooperative Primary Tests: 23B	X	X
4	5	10/72	Comprehensive Tests of Basic Skills: 2-R	X	X
1	5	9/72	Comprehensive Tests of Basic Skills: 2-Q	X	X
2	5	10/72	Comprehensive Tests of Basic Skills: 2-Q	X	X
4	5	10/71	Comprehensive Tests of Basic Skills: 2-R	X	X
		10/72	Comprehensive Tests of Basic Skills: 2-R	X	X
		9/73	Comprehensive Tests of Basic Skills: 2-R	X	X
1	5	9/72-5/73*	Stanford Diagnostic Reading Test: Comprehension	X	
		9/72-5/73*	Iowa Tests of Basic Skills: Arithmetic Skills		X
2	5	10/73	Iowa Tests of Educational Development: E	X	X
1	5	10/72-4/73*	Iowa Tests of Educational Development: F	X	X
1	6	10/71	Comprehensive Tests of Basic Skills: 2-Q	X	X
	5	9/72	Comprehensive Tests of Basic Skills: 2-Q	X	X
1	5/6	10/72	Comprehensive Tests of Basic Skills: 2-R	X	X
			Comprehensive Tests of Basic Skills: 2-Q	X	X
1	6	10/71	Comprehensive Tests of Basic Skills: 2-Q	X	X
1	6	10/71	Comprehensive Tests of Basic Skills: 2-Q	X	X
		10/72	Comprehensive Tests of Basic Skills: 2-Q	X	X

N=20

* Pre- and Post Tests